

Dormindo como uma pedra: Explorando fatores que moldam a qualidade do sono com modelos preditivos

Análise estatística da qualidade do sono

Henrique César Higino Holanda Cordeiro

Rafael do Nascimento Moura

Centro de Informática (CIn)
Universidade Federal de Pernambuco (UFPE)

Centro de Informática (CIn)
Universidade Federal de Pernambuco (UFPE)

Recife, Brasil
hchhc@cin.ufpe.br

Recife, Brasil
rnm4@cin.ufpe.br

Abstract—Este projeto propõe a criação de um modelo preditivo baseado em aprendizado profundo para analisar e prever a qualidade do sono utilizando dados comportamentais e fisiológicos. As atividades incluirão a aplicação de técnicas de aprendizado de máquina, com foco em modelos como Gradient Boosting Machine (GBM) e redes neurais perceptron multilaminadas (MLP). Os resultados visam identificar padrões relevantes e propor recomendações para melhorar a qualidade do sono, contribuindo para a saúde e qualidade de vida.

Index Terms—Sono, Cotidiano, Gradient Boosting Machine, GBM, Multilayer Perceptron, MLP, Saúde.

I. INTRODUÇÃO

O dia de qualquer pessoa sempre começa e termina da mesma forma, respectivamente, acordando e dormindo. Dormir é uma das ferramentas mais importantes para que a mente não sobrecarregue e o corpo funcione corretamente, tornando essa atividade tão essencial quanto se alimentar, beber água ou realizar exercícios. Segundo a Organização Mundial da Saúde (OMS), distúrbios do sono afetam cerca de 40% dos brasileiros, contribuindo para problemas como obesidade, ansiedade e doenças cardiovasculares.

No entanto, o estilo de vida moderno tem prejudicado a qualidade do sono de muitas pessoas. Fatores como excesso de exposição a telas antes de dormir, horários de trabalho irregulares e altos níveis de estresse diário estão entre os principais causadores desse problema. Monitorar e agir para que as condições de sono sejam as melhores possíveis é indispensável para a manutenção da saúde física e mental.

Este projeto busca desenvolver um modelo preditivo baseado em aprendizado profundo para analisar e prever a qualidade do sono com base em fatores comportamentais e fisiológicos. O objetivo é identificar padrões complexos que influenciam o descanso e fornecer insights práticos sobre as condições ideais para um sono reparador. Os resultados deste estudo podem contribuir significativamente para o de-

envolvimento de ferramentas que auxiliem no diagnóstico de distúrbios do sono e na criação de recomendações individuais, promovendo uma melhor qualidade de vida para os indivíduos.

II. OBJETIVO

Geral

- Desenvolver um modelo preditivo baseado em aprendizado profundo para análise e previsão da qualidade do sono com base em fatores comportamentais e fisiológicos.

Específicos

- Realizar uma análise exploratória dos dados para identificar padrões e relações entre os fatores comportamentais e fisiológicos que influenciam o sono.
- Implementar e treinar modelos de rede neural (GBM e MLP) para prever a qualidade do sono.
- Avaliar o desempenho dos modelos utilizando métricas como Erro Quadrático Médio (MSE) e coeficiente R^2 .
- Identificar os fatores mais relevantes que afetam a qualidade do sono com base nos resultados.
- Propor recomendações práticas para melhorar a qualidade do sono.

III. JUSTIFICATIVA

O estilo de vida, incluindo horários ativos, hábitos alimentares e atividades físicas, é essencial para a saúde do sono. No entanto, o cotidiano urbano frequentemente interfere nesse ciclo, prejudicando o descanso por fatores como jornadas irregulares e exposição a dispositivos eletrônicos.

Pesquisas sobre sono ainda exploram pouco o aprendizado profundo para identificar padrões complexos em dados comportamentais e fisiológicos. Este projeto utiliza um *dataset* estruturado e modelos preditivos para identificar fatores que afetam o sono, visando apoiar diagnósticos precoces de distúrbios e fornecer recomendações práticas para melhorar a qualidade de vida.

IV. METODOLOGIA

Este projeto será conduzido em etapas bem definidas para a criação do modelo preditivo. As etapas descritas a seguir detalham o processo a ser utilizado, desde o pré-processamento dos dados até a análise dos resultados.

A. Coleta e Descrição dos Dados

O dataset utilizado no projeto foi o *Health and Sleep Statistics*, que contém 12 *features* relacionadas a informações demográficas, comportamentais e fisiológicas dos indivíduos, incluindo:

- **ID do usuário:** identificação única.
- **Idade e gênero:** dados demográficos.
- **Qualidade do sono:** variável-alvo a ser predita.
- **Horários de dormir e acordar:** dados comportamentais.
- **Passos diários, calorias queimadas e nível de atividade física:** dados relacionados a atividade.
- **Hábitos alimentares, distúrbios do sono e uso de medicação:** informações fisiológicas e de saúde.

Esses dados estão fornecidos em formato tabular.

B. Pré-processamento dos Dados

Para garantir a qualidade dos dados e a eficácia do modelo, será realizado um pré-processamento com os seguintes passos:

- **Tratamento de valores ausentes:** Valores ausentes serão identificados e tratados, utilizando técnicas como imputação (média, moda ou mediana) dependendo da natureza da variável, caso existam.
- **Codificação de variáveis categóricas:** Dados como *gênero* e *hábitos alimentares* serão codificados usando *One-Hot Encoding* ou métodos nativos de modelos como o *CatBoost*.
- **Normalização:** Variáveis contínuas, como idade e passos diários, serão ajustadas para evitar escalas discrepantes entre os atributos.
- **Divisão dos dados:** O dataset será dividido em três subconjuntos:
 - **Treinamento:** 70% dos dados.
 - **Validação:** 15% dos dados.
 - **Teste:** 15% dos dados.
- **Análise exploratória dos dados (EDA):** Gráficos de dispersão, histogramas e matrizes de correlação serão gerados para identificar padrões e relações entre as variáveis.

C. Seleção do Modelo

Com base nas características do dataset, o modelo **GBM**, especificamente o *CatBoost*, como abordagem principal se mostra a melhor escolha devido à sua eficiência em dados tabulares e capacidade nativa de lidar com variáveis categóricas. Além disso, **MLP** será utilizado para comparação de desempenho, avaliando a adequação de abordagens baseadas em aprendizado profundo neste contexto.

D. Treinamento do Modelo

- **Configuração inicial:** Será configurado com hiperparâmetros padrão para os modelos.
- **Tuning de hiperparâmetros:** Otimização dos hiperparâmetros mais relevantes por *grid search*, incluindo:
 - Número de árvores.
 - Taxa de aprendizado.
 - Estratégias de regularização para evitar *overfitting*.
- **Validação cruzada:** Um esquema de validação cruzada com 5 *folds* será usado para avaliar a estabilidade e a robustez do modelo.

E. Avaliação do Modelo

A avaliação do desempenho do modelo seguirá as seguintes métricas:

- **Erro Quadrático Médio (RMSE):** Mede a diferença entre os valores preditos e reais da qualidade do sono.
- **Coefficiente de Determinação (R^2):** Avalia o grau de explicação das variáveis independentes em relação à variável-alvo.
- **Importância das *features*:** Identificar os fatores mais influentes na qualidade do sono.

F. Análise e Interpretação dos Resultados

Após o treinamento e a avaliação, os resultados serão interpretados para:

- Identificar os principais fatores que impactam a qualidade do sono.
- Comparar o desempenho entre o modelo principal (*GBM*) e a abordagem alternativa (*MLP*), destacando suas vantagens e limitações em dados comportamentais e fisiológico.
- Validar a eficácia do modelo preditivo e gerar recomendações práticas para a melhoria do sono.

G. Ferramentas Utilizadas

O projeto será implementado em **Python**, utilizando as seguintes bibliotecas e frameworks:

- **Pandas e NumPy:** Para manipulação e análise dos dados.
- **Matplotlib e/ou Seaborn:** Para visualização e análise exploratória.
- **CatBoost e Scikit-learn:** Para treinamento e avaliação dos modelos.
- **Jupyter Notebook:** Para desenvolvimento interativo e organização do código.

V. CRONOGRAMA DE ATIVIDADES

Atividades Programadas

Data	Atividade
22/01	Definir o Banco de Dados
26/01	Entregar Proposta de Projeto
29/01	Pré-processamento dos Dados
01/02	Analisar e Definir Relações
09/02	Confirmar Relações Relevantes e Equacionar
15/02	Começar implementação do Modelo
22/02	Fazer as Avaliações dos Resultados
01/03	Revisar e Preparar Apresentação
02/03	Escrever Relatório do Projeto
18/03	Entregar Relatório e Materiais do Projeto
19/03*	Apresentação do Projeto

**Datas sujeitas a modificação*

VI. REFERÊNCIAS

- [1] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, Dive into Deep Learning, release 0.17.1, 2023. [Online]. Available: <https://d2l.ai/>.
- [2] M. Nielsen, Neural Networks and Deep Learning. Dec. 2019. [Online]. Available: <http://neuralnetworksanddeeplearning.com/>.
- [3] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011. [Online]. Available: <https://scikit-learn.org/>.
- [4] H. Aksoy. Kaggle. “Health and Sleep Statistics Dataset,” Sep. 2024. [Online]. Available: <https://www.kaggle.com/datasets/hanksoy/health-and-sleep-statistics/data>.